

*Génét. Sél. Evol.*, 1983, 15 (2), 201-224

## Sire evaluation for ordered categorical data with a threshold model

D. GIANOLA\* and J.L. FOULLEY\*\*

\* *Department of Animal Science, University of Illinois, Urbana, Illinois 61801, U.S.A.*

\*\* *I.N.R.A., Station de Génétique quantitative et appliquée,  
Centre de Recherches Zootechniques, F 78350 Jouy-en-Josas.*

### Summary

A method of evaluation of ordered categorical responses is presented. The probability of response in a given category follows a normal integral with an argument dependent on fixed thresholds and random variables sampled from a conceptual distribution with known first and second moments, *a priori*. The prior distribution and the likelihood function are combined to yield the posterior density from which inferences are made. The mode of the posterior distribution is taken as an estimator of location. Finding this mode entails solving a non-linear system; estimation equations are presented. Relationships of the procedure to "generalized linear models" and "normal scores" are discussed. A numerical example involving sire evaluation for calving ease is used to illustrate the method.

*Key words* : sire evaluation, categorical data, threshold characters, Bayesian methods.

### Résumé

*Evaluation des reproducteurs sur un caractère discret ordonné,  
sous l'hypothèse d'un déterminisme continu sous-jacent à seuils*

Cet article présente une méthode d'évaluation des reproducteurs sur un caractère à expression discrète et ordonnée. La probabilité de réponse dans une catégorie donnée est exprimée comme l'intégrale d'une loi normale dont les bornes dépendent de seuils fixés et de variables aléatoires de premiers et deuxièmes moments connus. La distribution *a priori* des paramètres et la fonction de vraisemblance sont combinées en vue de l'obtention de la densité *a posteriori* qui sert de base à l'inférence statistique. Les paramètres sont estimés par les modes *a posteriori*, ce qui conduit à la résolution d'un système d'équations non linéaires. Les relations qui apparaissent entre cette méthode et celles du modèle linéaire généralisé d'une part, et des scores normaux d'autre part, sont discutées. Enfin, l'article présente une illustration numérique de cette méthode qui a trait à l'évaluation de taureaux sur les difficultés de naissance de leurs produits.

*Mots clés* : évaluation des reproducteurs, données discrètes, caractères à seuil, méthode Bayésienne.

## I. Introduction

Animal breeding data are often categorical in expression, i.e., the response variable being measured is an assignment into one of several mutually exclusive and exhaustive response categories. For example, litter size in sheep is scored as 0, 1, 2, 3 or more lambs born per ewe exposed to the ram or to artificial insemination in a given breeding season. The analysis may be directed to examine relationships between the categorical variate in question and a set of explanatory variables, to estimate functions and test hypotheses about parameters, to assess the relative importance of different sources of variation, or to rank a set of candidates for selection, i.e., sire or dam evaluation.

If the variable to be predicted, e.g., sire's genetic merit, and the data follow a multivariate normal distribution, best linear unbiased prediction (HENDERSON, 1973) is the method of choice ; a sire evaluation would be in this instance the maximum likelihood estimate of the best predictor. Categorical variates, however, are not normally distributed and linear methodology is difficult to justify as most of the assumptions required are clearly violated (THOMPSON, 1979 ; GIANOLA, 1982).

If the response variable is polychotomous, i.e., the number of response categories is larger than 2, it is essential to distinguish whether the categories are ordered or unordered. Perhaps with the exception of some dairy cattle type scoring systems, most polychotomous categorical variables of interest in animal breeding are ordered. In the case of litter size in sheep, for example, the response categories can be ordered along a fecundity gradient, i.e., from least prolific to most prolific. Quantitative geneticists have used the threshold model to relate a hypothetical, underlying continuous scale to the outward categorical responses (DEMPSTER and LERNER, 1950 ; FALCONER, 1965, 1967). With this model, it would be possible to score or scale response categories so as to conform with intervals of the normal distribution (KENDALL and STUART, 1961 ; SNELL, 1964 ; GIANOLA and NORTON, 1981) and then applying linear methods on the scaled data. One possible set of scores would be simple integers (HARVEY, 1982) although in most instances scores other than integers may be preferable (SNELL, 1964).

Additional complications arise in scaling categorical data in animal breeding. The error and the expectation structures of routinely used models are complex, and the methods of scaling described in the literature are not suitable under these conditions. For example, applications of Snell's scaling procedure to cattle data (TONG *et al.*, 1977 ; FERNANDO *et al.*, 1983) required "sires" to be regarded as a fixed set, as opposed to random samples from a conceptual population. Further, scaling alters the distribution of errors and changes in the variance-covariance structure need to be considered in the second stage of the analysis. Unfortunately, the literature does not offer guidance on how to proceed in this respect.

This paper presents a method of analyzing ordered categorical responses stemming from an underlying continuous scale where gene substitutions are made. The emphasis is on prediction of genetic merit in the underlying scale based on prior information about the population from which the candidates for selection are sampled. Relationships of the procedure with the extension of "generalized linear models" presented by THOMPSON (1979) and with the method of "normal scores" (KENDALL and STUART, 1961), are discussed. A small example with calving difficulty data is used to illustrate computational aspects.

II. Methodology

*Data.* The data are organized into an  $s \times m$  contingency table, where the  $s$  rows represent individuals or combinations of levels of explanatory variables, and the  $m$  columns indicate mutually exclusive and exhaustive ordered categories of response. The form of this array is presented in Table 1, where  $n_{jk}$  is the number of experimental units responding in the  $k^{\text{th}}$  category under the conditions of the  $j^{\text{th}}$  row. Row totals,  $n_{j\cdot}$  ( $j=1,\dots,s$ ), are assumed fixed, the only restriction being  $n_{j\cdot} \neq 0$  for all values of  $j$ . If the  $s$  rows represent individuals in which a polychotomous response is evaluated, then  $n_{j\cdot} = 1$ , for  $j=1,\dots,s$ . In fact, the requirement of non-null row totals can be relaxed since, as shown later, prior information can be used to predict the genetic merit of an individual "without data" in the contingency table from related individuals "with data". The random variables of interest are  $n_{j1}, n_{j2}, \dots, n_{jk}$ , for  $j=1,\dots,s$ . Since the marginal totals are fixed, the table can be exactly described by a model with  $s(m-1)$  parameters. However, a parsimonious model is desired.

TABLE 1

*Ordered categorical data arranged as an  $s \times m$  contingency table.*  
*Table de contingence  $s \times m$  figurant les données discrètes ordonnées.*

Row <sup>a</sup>	Category of response <sup>b</sup>				Fixed total <sup>c</sup>
	1	2 . . . k	. . . m		
1	$n_{11}$	$n_{12} \dots n_{1k}$	$\dots n_{1m}$	$n_{1.}$	
2	$n_{21}$	$n_{22} \dots n_{2k}$	$\dots n_{2m}$	$n_{2.}$	
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	
j	$n_{j1}$	$n_{j2} \dots n_{jk}$	$\dots n_{jm}$	$n_{j.}$	
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	
s	$n_{s1}$	$n_{s2} \dots n_{sk}$	$\dots n_{sm}$	$n_{s.}$	

a) A row represents an individual or a combination of levels of explanatory variables.

b)  $n_{jk}$  is the number of counts in the  $k^{\text{th}}$  category of response in the  $j^{\text{th}}$  row.

c)  $n_{j\cdot} = \sum_{k=1}^m n_{jk}$

The data in the contingency table can be represented symbolically by the  $m \times s$  matrix

$$Y = [Y_1, Y_2, \dots, Y_j, \dots, Y_s] \tag{1}$$

where  $Y_j$  is an  $m \times 1$  vector

$$Y_j = \sum_{r=1}^{n_j} Y_{jr} \quad (2)$$

and  $Y_{jr}$  is an  $m \times 1$  vector having a 1 in the row corresponding to the category of response of the  $j$ <sup>th</sup> experimental unit and zeroes elsewhere.

*Inferences.* The data  $Y$  are jointly distributed with a parameter vector  $\theta$ , the joint density being  $f(Y, \theta)$ . Inferences are based using Bayes theorem (LINDLEY, 1965).

$$f(\theta|Y) = g(Y|\theta) \cdot p(\theta)/t(Y) \quad (3)$$

where  $t(Y)$  is the marginal density of  $Y$ ;  $p(\theta)$  is the *a priori* density of  $\theta$ , which reflects the relative uncertainty about  $\theta$  before the data  $Y$  become available ;  $g(Y|\theta)$  is the likelihood function, and  $f(\theta|Y)$  is the *a posteriori* density. Since  $t(Y)$  does not vary with  $\theta$ , the posterior density can be written as

$$f(\theta|Y) \propto g(Y|\theta) \cdot p(\theta) \quad (4)$$

As BOX and TIAO (1973) have pointed out, all the information about  $\theta$ , after the data have been collected, is contained in  $f(\theta|Y)$ . If one could derive the posterior density, probability statements about  $\theta$  could be made, *a posteriori*, from  $f(\theta|Y)$ . However, if "realistic" functional forms are considered for  $p(\theta)$  or  $g(Y|\theta)$ , one cannot always obtain a mathematically tractable, integrable, expression for  $f(\theta|Y)$ .

In this paper, we characterize the posterior density with a point estimator, its mode. The mode is the function of the data which minimizes the expected posterior loss when the loss function is

$$\sum_i d_i, \text{ with } d_i \begin{cases} 1, \text{ for } |\hat{\theta}_i - \theta_i| > \epsilon \\ 0, \text{ otherwise} \end{cases}$$

where  $\epsilon$  is a positive but arbitrarily small number (PRATT, RAIFFA and SCHLAIFER, 1965). The mean and the median are the functions of the data which minimize expected posterior quadratic error loss and absolute error loss, respectively (FERGUSON, 1967). However,  $E(\theta|Y)$  and the posterior median are generally more difficult to compute than the posterior mode.

*Threshold model.* It is assumed that the response process is related to an underlying continuous variable,  $\ell$ , and to a set of fixed thresholds

$$\delta' = [\delta_1 < \delta_2, \dots < \delta_{m-1}] \quad (5)$$

with  $\delta_0 = -\infty$ , and  $\delta_m = \infty$ . The distribution of  $\ell$ , in the context of the multifactorial model of quantitative genetics, can be assumed normal (DEMPSTER and LERNER, 1950 ; CURNOW and SMITH, 1975 ; BULMER, 1980 ; GIANOLA, 1982) as this variate is regarded as the result of a linear combination of small effects stemming from alleles at a large number of loci, plus random environmental components. Associated with each row in the table, there is a location parameter  $\eta_j$ , so that the underlying variate for the  $q$ <sup>th</sup> experimental unit in the  $j$ <sup>th</sup> row can be written as

$$\ell_{jq} = \eta_j + \epsilon_{jq} \quad (6)$$

$j=1, \dots, s$  and  $q=1, \dots, n_j$ , and  $\varepsilon_{jq} \sim \text{IID } N(0, \sigma^2)$ , where IID stands for "independent and identically distributed". Further, the parameter  $\eta_j$  is given a linear structure

$$\eta_j = \mathbf{q}'_j \mathbf{v} + \mathbf{z}'_j \mathbf{u} \quad (7)$$

where  $\mathbf{q}'_j$  and  $\mathbf{z}'_j$  are known row vectors, and  $\mathbf{v}$  and  $\mathbf{u}$  are unknown vectors corresponding to fixed and random effects, respectively, in linear model analyses (e.g., SEARLE, 1971 ; HENDERSON, 1975). All location parameters in the contingency table can be written as

$$\boldsymbol{\eta} = \mathbf{Q}\mathbf{v} + \mathbf{Z}\mathbf{u} \quad (8)$$

where  $\boldsymbol{\eta}$  is of order  $s \times 1$ , and  $\mathbf{Q}$  and  $\mathbf{Z}$  are matrices of appropriate order, with  $\mathbf{v}$  defined such that  $\mathbf{Q}$  has full column rank  $r$ .

Given  $\eta_j$ , the probability of response in the  $k^{\text{th}}$  category under the conditions of the  $j^{\text{th}}$  row is

$$P_{jk} = \text{Prob}\{\delta_{k-1} < \ell < \delta_k | \eta_j\} = \Phi\left(\frac{\delta_k - \eta_j}{\sigma}\right) - \Phi\left(\frac{\delta_{k-1} - \eta_j}{\sigma}\right) \quad (9)$$

where  $\Phi(\cdot)$  is the standard normal distribution function. Since  $\sigma$  is not identifiable, it is taken as the unit of measurement, i.e.,  $\sigma=1$ . Write  $\mathbf{Q} = [\mathbf{1} \ \mathbf{X}]$  such that  $\text{rank}(\mathbf{X}) = r-1$  with  $\mathbf{1}$  being a vector of ones. Then

$$\boldsymbol{\eta} = \mathbf{1}\mathbf{v} + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$$

where  $\boldsymbol{\beta}$  is a vector of  $r-1$  elements, and

$$\delta_k - \eta_j = (\delta_k - v_1) - (\mathbf{x}'_j \boldsymbol{\beta} + \mathbf{z}'_j \mathbf{u}) = t_k - \mu_j; \quad k=1, \dots, m-1, \quad t_k > t_{k-1} \\ j=1, \dots, s$$

with  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ . Hence, the probabilities in (9) can be written as

$$P_{jk} = \Phi(t_k - \mu_j) - \Phi(t_{k-1} - \mu_j); \quad k=1, \dots, m-1 \\ j=1, \dots, s \quad (10)$$

Several authors (e.g., ASHTON, 1972 ; BOCK, 1975 ; GIANOLA and FOULLEY, 1982) have approximated the normal integral with a logistic function. Letting

$$t_k^* - \mu_j^* = \frac{\Pi}{\sqrt{3}}(t_k - \mu_j) \quad (11)$$

we have

$$\Phi(t_k - \mu_j) \doteq [1 + e^{-(t_k^* - \mu_j^*)}]^{-1} \quad (12)$$

It follows that

$$P_{jk} \doteq [1 + e^{-(t_k^* - \mu_j^*)}]^{-1} - [1 + e^{-(t_{k-1}^* - \mu_j^*)}]^{-1} \\ = c_{jk} - c_{j(k-1)} \quad (13)$$

For  $-5 < t_k - \mu_j < 5$ , the difference between (12) and  $\Phi(t_k - \mu_j)$  does not exceed .022 (JOHNSON and KOTZ, 1970). In this paper, formulae appropriate for both the normal and the logistic distributions are presented.

Irrespective of the functional form used to compute  $P_{jk}$ , it is clear from (10) or (13), that the distribution of response probabilities by category is a function of the distance

between  $\mu_j$  and the thresholds. For example, suppose we have two rows, with parameters  $\mu_1$  and  $\mu_2$ , and two categories with threshold  $t_1$ . Then, using (10)

$$\begin{aligned} P_{11} &= \Phi(t_1 - \mu_1), P_{12} = 1 - P_{11} \\ P_{21} &= \Phi(t_1 - \mu_2), P_{22} = 1 - P_{21} \end{aligned}$$

If  $\mu_1 < t_1 < \mu_2$ , it follows that  $P_{11} > P_{21}$  and, automatically,  $P_{12} < P_{22}$ .

*Parameter vector and prior distribution.* The vector of variables to be estimated is

$$\theta' = [t', \beta', u'] \quad (14)$$

*A priori*,  $t$ ,  $\beta$  and  $u$  are assumed to be independent, each sub-vector following a multivariate normal distribution. Hence

$$p(\theta) \propto p_1(t) \cdot p_2(\beta) \cdot p_3(u) \quad (15)$$

where  $p_1(t)$ ,  $p_2(\beta)$  and  $p_3(u)$  are the *a priori* densities of  $t$ ,  $\beta$  and  $u$ , respectively. Explicitly

$$t \sim N(\tau, \Omega)$$

$$\beta \sim N(\alpha, \Gamma)$$

$$u \sim N(0, G)$$

where  $\Omega$  and  $\Gamma$  are diagonal covariance matrices, and  $G$  is a non-singular covariance matrix. In genetic applications,  $u$  is generally a vector of additive genetic values or sire effects, so  $G$  is a function of additive relationships and of the coefficient of heritability.

Equation (15) can be written as

$$p(\theta) \propto \exp\left\{-\frac{1}{2}[(t-\tau)' \Omega^{-1}(t-\tau) + (\beta-\alpha)' \Gamma^{-1}(\beta-\alpha) + u' G^{-1} u]\right\} \quad (16)$$

It will be assumed that prior knowledge about  $t$  and  $\beta$  is vague, that is,  $\Omega = \infty$ , and  $\Gamma = \infty$ . This implies that  $p_1(t)$  and  $p_2(\beta)$  are locally uniform and that the posterior density does not depend upon  $\tau$  and  $\alpha$ . The equation (16) becomes

$$\ell n p(\theta) \propto -\frac{1}{2} u' G^{-1} u \quad (17)$$

*Likelihood function and posterior density.* Given  $\theta$ , it is assumed that the indicator variables in  $Y$  are conditionally independent, following a multinomial distribution with probabilities  $P_{j1}, \dots, P_{jk}, \dots, P_{jm}$ ;  $j=1, \dots, s$ . The log-likelihood is then

$$\ell n g(Y|\theta) \propto \sum_{j=1}^s \sum_{k=1}^m n_{jk} \ell n(P_{jk}) \quad (18)$$

From (4), the log of the posterior density is equal to the sum of (17), (18) and an additive constant  $C$

$$L(\theta) = \ell n f(\theta|Y) \propto \sum_{j=1}^s \sum_{k=1}^m n_{jk} \ell n(P_{jk}) - \frac{1}{2} u' G^{-1} u + C \quad (19)$$

### III. Implementation

As pointed out previously, we take as estimator of  $\theta$  the value which maximizes  $L(\theta)$ , i.e., the mode of the log-posterior density. This requires differentiating (19) with respect to  $\theta$ , setting the vector of first derivatives equal to zero and solving for  $\hat{\theta}$ . However,

$$\frac{\partial L(\theta)}{\partial \theta} = 0 \quad (20)$$

is not linear in  $\theta$  so an iterative solution is required. The method of Newton-Raphson (DAHLQUIST and BJORCK, 1974) consists of iterating with

$$\hat{\theta}^{[i]} = \hat{\theta}^{[i-1]} - \left[ \frac{\partial^2 L(\theta)}{\partial \theta \partial \theta'} \right]_{\theta = \hat{\theta}^{[i-1]}}^{-1} \cdot \left[ \frac{\partial L(\theta)}{\partial \theta} \right]_{\theta = \hat{\theta}^{[i-1]}} \quad (21)$$

where  $\hat{\theta}^{[i]}$  is an approximation to the value of  $\hat{\theta}$ , with the suffix in brackets indicating the iterate number. Starting with a trial value  $\hat{\theta}^{[0]}$ , the process yields a sequence of approximations  $\hat{\theta}^{[1]}, \hat{\theta}^{[2]}, \dots, \hat{\theta}^{[i]}$  and, under certain second order conditions,

$$\lim_{j \rightarrow \infty} \hat{\theta}^{[j]} = \hat{\theta}$$

In practice, iteration stops when  $\Delta^{[i]} = \hat{\theta}^{[i]} - \hat{\theta}^{[i-1]} < \epsilon$ , the latter being a vector of arbitrarily small numbers. In this paper, we work with

$$\left[ -\frac{\partial^2 L(\theta)}{\partial \theta \partial \theta'} \right]_{\theta = \hat{\theta}^{[i-1]}} \Delta^{[i]} = \left[ \frac{\partial L(\theta)}{\partial \theta} \right]_{\theta = \hat{\theta}^{[i-1]}} \quad (22)$$

*First derivatives.* The normal case is considered first. Some useful results are the following

$$\frac{\partial P_{jk}}{\partial (t_k - \mu_j)} = \frac{\partial}{\partial (t_k - \mu_j)} [\Phi(t_k - \mu_j) - \Phi(t_{k-1} - \mu_j)] = \phi(t_k - \mu_j) \quad (23a)$$

$$\frac{\partial P_{jk}}{\partial t_k} = \phi(t_k - \mu_j) \quad (23b)$$

$$\frac{\partial P_{jk}}{\partial \mu_j} = \phi(t_{k-1} - \mu_j) - \phi(t_k - \mu_j) \quad (23c)$$

$$\frac{\partial P_{jk}}{\partial \beta} = [\phi(t_{k-1} - \mu_j) - \phi(t_k - \mu_j)] x_j \quad (23d)$$

with  $z_j$  replacing  $x_j$  in the derivative of  $P_{jk}$  with respect to  $u$ . Then

$$\frac{\partial L(\theta)}{\partial t_k} = \sum_{j=1}^s \left[ \frac{n_{jk}}{P_{jk}} - \frac{n_{j(k+1)}}{P_{j(k+1)}} \right] \phi(t_k - \mu_j) \quad (24)$$

$k = 1, \dots, m-1$

$$\frac{\partial L(\theta)}{\partial \beta} = \sum_{j=1}^s \left\{ \sum_{k=1}^m n_{jk} \left[ \frac{\phi(t_{k-1} - \mu_j) - \phi(t_k - \mu_j)}{P_{jk}} \right] \right\} x_j \quad (25)$$

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \mathbf{u}} = \sum_{j=1}^s \left\{ \sum_{k=1}^m n_{jk} \left[ \frac{\phi(t_{k-1}-\mu_j) - \phi(t_k-\mu_j)}{P_{jk}} \right] \right\} \mathbf{z}_j - \mathbf{G}^{-1} \mathbf{u} \quad (26)$$

Letting

$$v_j = \sum_{k=1}^m n_{jk} \left[ \frac{\phi(t_{k-1}-\mu_j) - \phi(t_k-\mu_j)}{P_{jk}} \right]$$

and  $\mathbf{v}' = [v_1, \dots, v_j, \dots, v_s]$ , (25) and (26) can be written as

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \mathbf{X}' \mathbf{v} \quad (27)$$

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \mathbf{u}} = \mathbf{Z}' \mathbf{v} - \mathbf{G}^{-1} \mathbf{u} \quad (28)$$

If the logistic function is used to approximate the normal

$$\frac{\partial P_{jk}}{\partial t_k^*} = c_{jk}(1-c_{jk}) \quad (29a)$$

$$\frac{\partial P_{jk}}{\partial \boldsymbol{\beta}^*} = [c_{j(k-1)}(1-c_{j(k-1)}) - c_{jk}(1-c_{jk})] \mathbf{x}_j \quad (29b)$$

and the equivalents of (24), (27) and (28) are

$$\frac{\partial L(\boldsymbol{\theta})}{\partial t_k^*} = \sum_{j=1}^s \left[ \frac{n_{jk}}{P_{jk}} - \frac{n_{j(k+1)}}{P_{j(k+1)}} \right] c_{jk} (1-c_{jk}) \quad (30)$$

$k = 1, \dots, m-1$

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}^*} = \mathbf{X}' \mathbf{v}^* \quad (31)$$

and

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \mathbf{u}^*} = \mathbf{Z}' \mathbf{v}^* - \mathbf{G}^{-1} \mathbf{u}^* (\Pi^2/3) \quad (32)$$

where  $\mathbf{v}^*$  is a  $s \times 1$  vector with typical element

$$v_j^* = \sum_{k=1}^m n_{jk} \left\{ \frac{c_{j(k-1)}[1-c_{j(k-1)}] - c_{jk}(1-c_{jk})}{P_{jk}} \right\}$$

Note that  $c_{jk}(1-c_{jk})$  in the logistic case replaces  $\phi(t_k-\mu_j)$  which appears when the normal distribution is used.

*Second derivatives.* The following derivatives need to be considered : a) threshold : threshold ; b) threshold :  $\boldsymbol{\beta}$  ; c) threshold :  $\mathbf{u}$  ; d)  $\boldsymbol{\beta} : \boldsymbol{\beta}'$  ; e)  $\boldsymbol{\beta} : \mathbf{u}'$ , and f)  $\mathbf{u} : \mathbf{u}'$ .

a) In the threshold : threshold derivatives, we start by writing

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial t_g \partial t_k} = \frac{\partial}{\partial t_g} \left\{ \sum_{j=1}^s \left[ \frac{n_{jk}}{P_{jk}} - \frac{n_{j(k+1)}}{P_{j(k+1)}} \right] \frac{\partial P_{jk}}{\partial t_k} \right\} \quad (33)$$



which holds both in the normal and logistic case (see equations 23b, 24 and 29a and 30). After algebra

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta})}{\partial t_g \partial t_k} &= \sum_{j=1}^S \left\{ \left[ \frac{n_{jk}}{P_{jk}} - \frac{n_{j(k+1)}}{P_{j(k+1)}} \right] \frac{\partial^2 P_{jk}}{\partial t_g \partial t_k} \right. \\ &\quad \left. + \left[ \frac{n_{j(k+1)}}{P_{j(k+1)}^2} \frac{\partial P_{j(k+1)}}{\partial t_g} - \frac{n_{jk}}{P_{jk}^2} \frac{\partial P_{jk}}{\partial t_g} \right] \frac{\partial P_{jk}}{\partial t_k} \right\} \end{aligned} \quad (34)$$

Considerable simplification is obtained by replacing  $n_{jk}$  by  $E(n_{jk}|\boldsymbol{\theta}) = n_j P_{jk}$ . Equation (34) becomes

$$E\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial t_g \partial t_k}\right] = \sum_{j=1}^S n_j \left[ \frac{1}{P_{j(k+1)}} \cdot \frac{\partial P_{j(k+1)}}{\partial t_g} - \frac{1}{P_{jk}} \cdot \frac{\partial P_{jk}}{\partial t_g} \right] \frac{\partial P_{jk}}{\partial t_k} \quad (35)$$

When  $g = k$ , (35) becomes

$$E\left[\frac{\partial^2 L(\boldsymbol{\theta})}{(\partial t_k)^2}\right] = - \sum_{j=1}^S n_j \left[ \frac{1}{P_{j(k+1)}} + \frac{1}{P_{jk}} \right] \left( \frac{\partial P_{jk}}{\partial t_k} \right)^2 \quad (36)$$

In the normal case,

$$E\left[\frac{\partial^2 L(\boldsymbol{\theta})}{(\partial t_k)^2}\right] = \sum_{j=1}^S n_j \left[ \frac{P_{jk} + P_{j(k+1)}}{P_{jk} P_{j(k+1)}} \right] \Phi^2(t_k - \mu_j) \quad (37)$$

$k=1, \dots, m-1$

and when the logistic function is used

$$E\left[\frac{\partial^2 L(\boldsymbol{\theta})}{(\partial t_k^*)^2}\right] = \sum_{j=1}^S n_j \left[ \frac{P_{jk} + P_{j(k+1)}}{P_{jk} P_{j(k+1)}} \right] c_{jk}^2 (1 - c_{jk})^2 \quad (38)$$

$k=1, \dots, m-1$

When  $g = k+1$ , equation (35) in the normal case becomes

$$E\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial t_{k+1} \partial t_k}\right] = \sum_{j=1}^S n_j \frac{\Phi(t_{k+1} - \mu_j) \Phi(t_k - \mu_j)}{P_{j(k+1)}} \quad (39)$$

$k=1, \dots, m-1$

and in the logistic approximation

$$E\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial t_{k+1}^* \partial t_k^*}\right] = \sum_{j=1}^S n_j \frac{c_{j(k+1)} [1 - c_{j(k+1)}] c_{jk} (1 - c_{jk})}{P_{j(k+1)}} \quad (40)$$

$k=1, \dots, m-1$

Elsewhere, when  $|g-k| > 1$

$$E\left[\frac{\partial^2 \ell n L(\boldsymbol{\theta})}{\partial t_g \partial t_k}\right] = E\left[\frac{\partial^2 \ell n L(\boldsymbol{\theta})}{\partial t_g^* \partial t_k^*}\right] = 0 \quad (41)$$

b) To obtain the threshold :  $\beta$  derivatives, first write for the normal case

$$\frac{\partial^2 L(\theta)}{\partial \beta \partial t_k} = \frac{\partial}{\partial \beta} \left\{ \sum_{j=1}^s \left[ \frac{n_{jk}}{P_{jk}} - \frac{n_{j(k+1)}}{P_{j(k+1)}} \right] \phi(t_k - \mu_j) \right\}; k=1, \dots, m-1 \quad (42)$$

After algebra, and replacing  $n_{jk}$  by  $n_j P_{jk}$ , one obtains

$$E\left[\frac{\partial^2 L(\theta)}{\partial \beta \partial t_k}\right] = \sum_{j=1}^s n_j \cdot \phi(t_k - \mu_j) \left[ \frac{\phi(t_k - \mu_j) - \phi(t_{k-1} - \mu_j)}{P_{jk}} - \frac{\phi(t_{k+1} - \mu_j) - \phi(t_k - \mu_j)}{P_{j(k+1)}} \right] x_j \quad (43)$$

$k=1, \dots, m-1$

Now, letting

$$\ell(k, j) = -n_j \cdot \phi(t_k - \mu_j) \left[ \frac{\phi(t_k - \mu_j) - \phi(t_{k-1} - \mu_j)}{P_{jk}} - \frac{\phi(t_{k+1} - \mu_j) - \phi(t_k - \mu_j)}{P_{j(k+1)}} \right]$$

equation (43) can be written as

$$E\left[\frac{\partial^2 L(\theta)}{\partial \beta \partial t_k}\right] = - \sum_{j=1}^s \ell(k, j) x_j = -X' \ell(k) \quad (44)$$

$k=1, \dots, m-1$

where  $\ell(k)$  is a  $s \times 1$  vector with typical element  $\ell(k, j)$ . In the logistic case, we use  $\ell^*(k)$  and  $\ell^*(k, j)$ , with  $c_{jk}(1-c_{jk})$  instead of  $\phi(t_k - \mu_j)$ .

c) The threshold :  $u$  expected second derivatives are

$$E\left[\frac{\partial^2 L(\theta)}{\partial u \partial t_k}\right] = - \sum_{j=1}^s \ell(k, j) z_j = -Z' \ell(k) \quad (45)$$

$k=1, \dots, m-1$

with  $\ell^*(k)$  replacing  $\ell(k)$  in the logistic case.

d) To obtain the second partial derivatives with respect to  $\beta$ , write

$$\frac{\partial^2 L(\theta)}{\partial \beta \partial \beta'} = \frac{\partial}{\partial \beta'} \left\{ \sum_{j=1}^s \sum_{k=1}^m n_{jk} \left[ \frac{\phi(t_{k-1} - \mu_j) - \phi(t_k - \mu_j)}{P_{jk}} \right] x_j \right\} \quad (46)$$

which, after algebra, becomes

$$\begin{aligned} \frac{\partial^2 L(\theta)}{\partial \beta \partial \beta'} &= \sum_{j=1}^s \sum_{k=1}^m \left\{ \frac{n_{jk}}{P_{jk}} [\phi(t_{k-1} - \mu_j)(t_{k-1} - \mu_j) - \phi(t_k - \mu_j)(t_k - \mu_j)] \right. \\ &\quad \left. - \frac{n_{jk}}{P_{jk}^2} [\phi(t_{k-1} - \mu_j) - \phi(t_k - \mu_j)]^2 \right\} x_j x_j' \end{aligned} \quad (47)$$

Replacing  $n_{jk}$  by  $n_j P_{jk}$ , allows us to sum the first term of (47) over the index  $k$ . However,

$$\sum_{k=1}^m [\phi(t_{k-1} - \mu_j)(t_{k-1} - \mu_j) - \phi(t_k - \mu_j)(t_k - \mu_j)] = \phi(t_0 - \mu_j)(t_0 - \mu_j) - \phi(t_m - \mu_j)(t_m - \mu_j) = 0$$

so (47) can be written as

$$E\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right] = -\sum_{j=1}^s \sum_{k=1}^m n_{jk} \frac{[\phi(t_{k-1}-\mu_j) - \phi(t_k-\mu_j)]^2}{P_{jk}} \mathbf{x}_j \mathbf{x}_j' = -\mathbf{X}' \mathbf{W} \mathbf{X} \quad (48)$$

where  $\mathbf{W}$  is a diagonal matrix with typical element

$$w_{jj} = n_{jk} \frac{[\phi(t_{k-1}-\mu_j) - \phi(t_k-\mu_j)]^2}{P_{jk}} \quad (49)$$

When the logistic distribution is used,  $c_{jk}(1-c_{jk})$  replaces  $\phi(t_k-\mu_j)$  in (49), and the matrix of weights becomes  $\mathbf{W}^*$ .

e) The  $\boldsymbol{\beta} : \mathbf{u}'$  derivatives are

$$E\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \mathbf{u}'}\right] = -\mathbf{X}' \mathbf{W} \mathbf{Z} \quad (50)$$

f) Similarly, the  $\mathbf{u} : \mathbf{u}'$  derivatives are

$$E\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \mathbf{u} \partial \mathbf{u}'}\right] = -\mathbf{Z}' \mathbf{W} \mathbf{Z} - \mathbf{G}^{-1} \quad (51)$$

*Estimation equations.* The first and second derivatives of the previous sections are then used in (22). The algorithm becomes a "scoring" procedure as expected second derivatives are utilized, and (22) can be written as :

$$\begin{bmatrix} \mathbf{T}^{[i-1]} & \mathbf{L}'^{[i-1]} \mathbf{X} & \mathbf{L}^{[i-1]} \mathbf{Z} \\ \mathbf{X}' \mathbf{L}^{[i-1]} & \mathbf{X}' \mathbf{W}^{[i-1]} \mathbf{X} & \mathbf{X}' \mathbf{W}^{[i-1]} \mathbf{Z} \\ \mathbf{Z}' \mathbf{L}^{[i-1]} & \mathbf{Z}' \mathbf{W}^{[i-1]} \mathbf{X} & \mathbf{Z}' \mathbf{W}^{[i-1]} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \Delta_t^{[i]} \\ \Delta_{\boldsymbol{\beta}}^{[i]} \\ \Delta_{\mathbf{u}}^{[i]} \end{bmatrix} = \begin{bmatrix} \mathbf{p}^{[i-1]} \\ \mathbf{X}' \mathbf{v}^{[i-1]} \\ \mathbf{Z}' \mathbf{v}^{[i-1]} - \mathbf{G}^{-1} \mathbf{u}^{[i-1]} \end{bmatrix} \quad (52)$$

where :

i)  $\mathbf{T}^{[i-1]}$  is an  $(m-1) \times (m-1)$  banded matrix with elements equal to the negatives of (37) or (38) in the diagonal ; (39) and (40) with negative signs in  $k, k+1$  or  $k+1, k$  off-diagonals ( $k=1, \dots, m-1$ ), and zeroes elsewhere. For example, if the number of response categories is 3, in the normal case we have, neglecting suffixes :

$$\mathbf{T} = \begin{bmatrix} \sum_{j=1}^s n_{j1} \left[ \frac{P_{j1}+P_{j2}}{P_{j1} P_{j2}} \right] \phi^2(t_1-\mu_j) - \sum_{j=1}^s n_{j1} \frac{\phi(t_1-\mu_j)\phi(t_2-\mu_j)}{P_{j2}} & \sum_{j=1}^s n_{j1} \left[ \frac{P_{j2}+P_{j3}}{P_{j2} P_{j3}} \right] \phi^2(t_2-\mu_j) \\ - \sum_{j=1}^s n_{j1} \frac{\phi(t_2-\mu_j)\phi(t_1-\mu_j)}{P_{j2}} & \end{bmatrix}$$

$$\text{ii) } \mathbf{X}' \mathbf{L} = [\mathbf{X}' \ell(1) \quad \mathbf{X}' \ell(2) \quad \dots \quad \mathbf{X}' \ell(m-1)],$$

$$\mathbf{Z}' \mathbf{L} = [\mathbf{Z}' \ell(1) \quad \mathbf{Z}' \ell(2) \quad \dots \quad \mathbf{Z}' \ell(m-1)],$$

with  $\ell(k)$  as in (44), or  $\ell^*(k)$  in the logistic case, and

$$\text{iii) } \mathbf{p}^{[i-1]} = \left\{ \sum_{j=1}^s \left[ \frac{n_{jk}}{P_{jk}} - \frac{n_{j(k+1)}}{P_{j(k+1)}} \right] \phi(t_k-\mu_j) \right\}; k = 1, \dots, m-1$$

is an  $(m-1) \times 1$  vector.

Computations proceed by iteration with (52) starting with trial values  $\mathbf{t}^{[0]}$ ,  $\boldsymbol{\beta}^{[0]}$  and  $\mathbf{u}^{[0]}$  and stopping when certain rule is satisfied. Some possible trial values are discussed in the numerical example.

#### IV. Posterior inference

The exact functional form of the posterior density,  $f(\boldsymbol{\theta}|\mathbf{Y})$  (equation 19), is not known, and we have chosen to use the mode of this density as an estimator of  $\boldsymbol{\theta}$ . For breeding purposes, one might be interested in functions of  $\boldsymbol{\theta}$  rather than in  $\boldsymbol{\theta}$  itself. For example, we may wish to estimate a linear combination of response probabilities associated with a particular set of conditions in the contingency table :

$$f(\boldsymbol{\theta}) = \sum_{k=1}^m a_k P_{jk} = \sum_{k=1}^m a_j [(\Phi(t_k - \mu_j) - \Phi(t_{k-1} - \mu_j))]$$

where  $a_k$  is an arbitrary weight. One possible estimator of  $f(\boldsymbol{\theta})$  would be

$$f(\hat{\boldsymbol{\theta}}) = \sum_{k=1}^m a_j [\Phi(t_k - \hat{\mu}_j) - \Phi(t_{k-1} - \hat{\mu}_j)]$$

with  $\mathbf{t}$ ,  $\boldsymbol{\beta}$  and  $\mathbf{u}$  estimated from (52), and  $\hat{\mu}_j$  calculated from  $\hat{\boldsymbol{\beta}}$  and  $\hat{\mathbf{u}}$ . While  $\hat{\boldsymbol{\theta}}$  is the mode of the posterior density of  $\boldsymbol{\theta}$ , it does not follow that  $f(\hat{\boldsymbol{\theta}})$  is the mode of the posterior of  $f(\boldsymbol{\theta})$ . Only the median of the posterior is known to have this desirable invariance property (DE GROOT, 1982). However, if the posterior is symmetric and unimodal, the median, the mode and the mean of this density will be the same.

COX and HINKLEY (1974) have shown that the posterior density is asymptotically normal, and LEONARD (1972) has pointed out, in the case of binary responses, that a normal density approximation to the posterior can be justified when none of the  $n_{jk}$  values are small. Hence, asymptotically

$$E[\mathbf{k}'\boldsymbol{\theta}|\mathbf{Y}] = \mathbf{k}'\hat{\boldsymbol{\theta}}, \quad (53)$$

$$\text{Var}[\mathbf{k}'\boldsymbol{\theta}|\mathbf{Y}] = \mathbf{k}'\mathbf{C}\mathbf{k}, \quad (54)$$

and

$$(\mathbf{k}'\boldsymbol{\theta} - \mathbf{k}'\hat{\boldsymbol{\theta}}) / [\mathbf{k}'\mathbf{C}\mathbf{k}]^{.5} \sim N(0,1) \quad (55)$$

where  $\mathbf{k}'$  is a vector of arbitrary constants,  $\hat{\boldsymbol{\theta}}$  is the mode of the posterior density, and  $\mathbf{C}$  is the inverse of the coefficient matrix in (52). Specifically,

$$\mathbf{C} = \{E[-\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}]\}^{-1}$$

which can be evaluated at  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ . Note that (55) permits probability statements, *a posteriori* and asymptotically, about linear combinations of  $\boldsymbol{\theta}$ .

Since the median, the mode and the mean are asymptotically the same,  $f(\mathbf{k}'\hat{\boldsymbol{\theta}})$  can be

justified by the invariance property of the median, as an estimator of  $f(\mathbf{k}'\boldsymbol{\theta})$  under absolute error loss. The posterior dispersion of  $f(\mathbf{k}'\boldsymbol{\theta})$  can then be approximated as

$$\text{Var} [f(\mathbf{k}'\boldsymbol{\theta})|\mathbf{Y}] \doteq \left\{ \frac{\partial [f(\mathbf{k}'\boldsymbol{\theta})]}{\partial (\mathbf{k}'\boldsymbol{\theta})} \right\}^2 \cdot (\mathbf{k}'\mathbf{C}\mathbf{k}) \quad (56)$$

### V. Evaluation of individuals without data in the contingency table

As pointed out by HENDERSON (1977), a common problem arising in animal breeding is the one where it is wished to evaluate the genetic merit of individuals without records from data contributed by related candidates. In the context of this paper, this is tantamount for obtaining an evaluation of individuals without entries in the  $s \times m$  contingency table (Table 1).

Let  $\mathbf{u}' = [\mathbf{u}'_1, \mathbf{u}'_2]$  represent a vector of additive genetic values in the underlying scale. The individuals in  $\mathbf{u}_1$  contribute data to the contingency table while those in  $\mathbf{u}_2$  do not. Assume, *a priori*, that

$$\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} \sim N \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{pmatrix} \mid k \right) \quad (57)$$

where  $\mathbf{G}_{ij}$  is the matrix of additive relationships between  $\mathbf{u}_i$  and  $\mathbf{u}_j$ , and  $k$  is a known scalar, the additive genetic variance in the underlying scale. Since the residual variance is assumed to be equal to one,  $k = h^2 / (1-h^2)$ , where  $h^2$  is heritability in the narrow sense.

As in (19), the log-posterior density is

$$L(\boldsymbol{\theta}) \propto \sum_{j=1}^s \sum_{k=1}^m n_{jk} \ell n (P_{jk}) - \frac{1}{2} (\mathbf{u}'_1 \mathbf{G}^{11} \mathbf{u}_1 + \mathbf{u}'_1 \mathbf{G}^{12} \mathbf{u}_2 + \mathbf{u}'_2 \mathbf{G}^{21} \mathbf{u}_1 + \mathbf{u}'_2 \mathbf{G}^{22} \mathbf{u}_2) \quad (58)$$

where

$$\begin{pmatrix} \mathbf{G}^{11} & \mathbf{G}^{12} \\ \mathbf{G}^{21} & \mathbf{G}^{22} \end{pmatrix} = \begin{pmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{pmatrix}^{-1} \quad (59)$$

The mode of the posterior density can then be obtained by iteration with

$$\begin{pmatrix} \mathbf{T}^{[i-1]} & \mathbf{L}'^{[i-1]}\mathbf{X} & \mathbf{L}'^{[i-1]}\mathbf{Z} & \mathbf{0} \\ \mathbf{X}'\mathbf{L}^{[i-1]} & \mathbf{X}'\mathbf{W}^{[i-1]}\mathbf{X} & \mathbf{X}'\mathbf{W}^{[i-1]}\mathbf{Z} & \mathbf{0} \\ \mathbf{Z}'\mathbf{L}^{[i-1]} & \mathbf{Z}'\mathbf{W}^{[i-1]}\mathbf{X} & \mathbf{Z}'\mathbf{W}^{[i-1]}\mathbf{Z} + \mathbf{G}^{11} & \mathbf{G}^{12} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}^{21} & \mathbf{G}^{22} \end{pmatrix} \begin{pmatrix} \Delta_t^{[i]} \\ \Delta_\beta^{[i]} \\ \Delta_{u_1}^{[i]} \\ \Delta_{u_2}^{[i]} \end{pmatrix} = \begin{pmatrix} \mathbf{P}^{[i-1]} \\ \mathbf{X}'\mathbf{v}^{[i-1]} \\ \mathbf{Z}'\mathbf{v}^{[i-1]} \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{G}^{11} & \mathbf{G}^{12} \\ \mathbf{G}^{21} & \mathbf{G}^{22} \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} \quad (60)$$

## VI. Relationship to "generalized" linear models

In a discussion of sire evaluation methods, THOMPSON (1979) pointed out some theoretical limitations of linear models for "all or none" variables. He suggested as an alternative to regard the data as binomially distributed with mean value  $\Phi(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})$ . In this setting, maximum likelihood estimates of  $\boldsymbol{\beta}$  and  $\mathbf{u}$  could be obtained iteratively from a set of equations similar to weighted least-squares, with the data vector  $\mathbf{y}$  replaced by  $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}[\mathbf{Y} - \Phi(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})]$ , where  $\mathbf{W}$  is diagonal, and with an also diagonal matrix of weights replacing the residual covariance matrix. This was interpreted by THOMPSON as a "generalized" linear model, in the sense of NELDER and WEDDERBURN (1972), in which  $\boldsymbol{\beta}$  and  $\mathbf{u}$  are regarded as constants. If  $\mathbf{u}$  is a vector of realized values of random variables instead of constants, THOMPSON said that it would be intuitively appealing to modify these "generalized" linear model equations in the same way as weighted least-squares equations are amended to obtain "mixed model" equations (HENDERSON, 1973). It turns out that the methodology developed in the present paper and in GIANOLA and FOULLEY (1982), yields the equations anticipated by THOMPSON.

Consider the binary case, i.e.,  $m=2$  so there is one threshold,  $t$ . As the value of the threshold by itself is of no interest, one can write the probability of response in the first category for the  $j^{\text{th}}$  row as

$$P_{j1} = \Phi(t - \mu_j) = \Phi(\gamma_j)$$

and then structure  $\boldsymbol{\gamma}' = [\gamma_1, \dots, \gamma_s]$  as

$$\boldsymbol{\gamma} = \mathbf{K}\boldsymbol{\lambda} + \mathbf{H}\mathbf{h}$$

where, *a priori*,  $\mathbf{h} \sim \mathbf{N}(\mathbf{0}, \mathbf{G})$ . In this case, as shown by GIANOLA and FOULLEY (1982), equations (52) become

$$\begin{bmatrix} \mathbf{K}'\mathbf{W}^{[i-1]}\mathbf{K} & \mathbf{K}'\mathbf{W}^{[i-1]}\mathbf{H} \\ \mathbf{H}'\mathbf{W}^{[i-1]}\mathbf{K} & \mathbf{H}'\mathbf{W}^{[i-1]}\mathbf{H} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \Delta_{\boldsymbol{\lambda}}^{[i]} \\ \Delta_{\mathbf{h}}^{[i]} \end{bmatrix} = \begin{bmatrix} \mathbf{K}'\mathbf{v}^{[i-1]} \\ \mathbf{H}'\mathbf{v}^{[i-1]} - \mathbf{G}^{-1}\mathbf{h}^{[i-1]} \end{bmatrix} \quad (61)$$

Now, adding the coefficient matrix in (61) postmultiplied by  $[\boldsymbol{\lambda}'^{[i-1]}\mathbf{h}'^{[i-1]}]'$  to both sides of the equation, and rearranging, one obtains

$$\begin{bmatrix} \mathbf{K}'\mathbf{W}^{[i-1]}\mathbf{K} & \mathbf{K}'\mathbf{W}^{[i-1]}\mathbf{H} \\ \mathbf{H}'\mathbf{W}^{[i-1]}\mathbf{K} & \mathbf{H}'\mathbf{W}^{[i-1]}\mathbf{H} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}^{[i]} \\ \mathbf{h}^{[i]} \end{bmatrix} = \begin{bmatrix} \mathbf{K}'\mathbf{W}^{[i-1]}\{\mathbf{K}\boldsymbol{\lambda}^{[i-1]} + \mathbf{H}\mathbf{h}^{[i-1]} + (\mathbf{W}^{[i-1]})^{-1}\mathbf{v}^{[i-1]}\} \\ \mathbf{H}'\mathbf{W}^{[i-1]}\{\mathbf{K}\boldsymbol{\lambda}^{[i-1]} + \mathbf{H}\mathbf{h}^{[i-1]} + (\mathbf{W}^{[i-1]})^{-1}\mathbf{v}^{[i-1]}\} \end{bmatrix} \quad (62)$$

This parallels a set of "mixed model" equations with residual covariance matrix replaced by the inverse of  $\mathbf{W}^{[i-1]}$ , and where the data are replaced by a "working variate"

$$\mathbf{K}\boldsymbol{\lambda}^{[i-1]} + \mathbf{H}\mathbf{h}^{[i-1]} + (\mathbf{W}^{[i-1]})^{-1}\mathbf{v}^{[i-1]}$$

which has a linear structure plus a residual  $(\mathbf{W}^{[i-1]})^{-1}\mathbf{v}^{[i-1]}$ .

## VII. Relationship to "normal scores"

Ordered categorical data are often analyzed by imposing a metric on the categories of the table. However, it is not always clear how scores should be assigned or computed. If the response process is assumed to be related to an underlying continuous variate, the

simplest scoring system would be the set of natural numbers  $\{1, 2, \dots, m\}$ . HARVEY (1982), suggested using this system and analyzing the scaled data with linear least-squares. However, it is difficult to justify linear models for categorical data (THOMPSON, 1979 ; GIANOLA, 1982). To illustrate this difficulty, suppose that a polychotomous response is examined, and that the response process is related to an underlying normal distribution. With  $m$  categories of response, the random variable of interest for the  $r^{\text{th}}$  experimental unit in the  $j^{\text{th}}$  condition is an  $m \times 1$  vector,  $\mathbf{v}_{jr}$ , containing a 1 in the category of response, and zeroes elsewhere. If  $\mathbf{a}' = [1, 2, \dots, m]$  is the set of the first  $m$  natural numbers and the response is in the  $k^{\text{th}}$  category, then

$$E(\mathbf{a}'\mathbf{v}_{jr}|\mu_j) = \sum_{k=1}^m a_k [\Phi(t_k - \mu_j) - \Phi(t_{k-1} - \mu_j)] \quad (63)$$

where  $\mu_j = \mathbf{x}'_j\boldsymbol{\beta} + \mathbf{z}'_j\mathbf{u}$ . However,

$$\frac{\partial E(\mathbf{a}'\mathbf{v}_{jr}|\mu_j)}{\partial \mu_j} = \sum_{k=1}^m a_k [\phi(t_{k-1} - \mu_j) - \phi(t_k - \mu_j)] \quad (64)$$

so a model linear in  $\mu_j$  cannot be justified as the effect of changes in the explanatory variables is not constant throughout the range of  $\mu_j$ . In addition,

$$\text{Var}(\mathbf{a}'\mathbf{v}_{jr}|\mu_j) = \sum_{k=1}^m a_k^2 [\Phi(t_k - \mu_j) - \Phi(t_{k-1} - \mu_j)] - \left\{ \sum_{k=1}^m a_k [\Phi(t_k - \mu_j) - \Phi(t_{k-1} - \mu_j)] \right\}^2 \quad (65)$$

is not constant as the variance depends on  $\mu_j$ .

KENDALL and STUART (1961) discuss "normal scores", a method in which the scores of  $m$  ordered response categories are developed in connection with an hypothetical, normally distributed variable  $y$  with mean  $\mu_j$  and variance 1. The model for the underlying variate would then be

$$y_{jr} = \mu_j + \varepsilon_{jr}$$

where  $\varepsilon_{jr} \sim N(0, 1)$ . Now, if the response is in the  $k^{\text{th}}$  category, using the notation of this paper :

$$\begin{aligned} E(y_{jr}|\mu_j, t_{k-1} < y < t_k) &= \mu_j + E(\varepsilon_{jr}|t_{k-1} - \mu_j < \varepsilon_{jr} < t_k - \mu_j) \\ &= \mu_j + \int_{t_{k-1} - \mu_j}^{t_k - \mu_j} \varepsilon f(\varepsilon) d\varepsilon / [\Phi(t_k - \mu_j) - \Phi(t_{k-1} - \mu_j)] \\ &= \mu_j + [\Phi(t_{k-1} - \mu_j) - \Phi(t_k - \mu_j)] / P_{jk} \end{aligned} \quad (66)$$

Since KENDALL and STUART have considered a single population, i.e., only one  $\mu$ , their "normal score" is the residue in (66). Further

$$\text{Var}(y_{jr}|\mu_j, t_{k-1} < y < t_k) = E[y_{jr}^2|\mu_j, t_{k-1} < y < t_k] - \{\mu_j + [\Phi(t_{k-1} - \mu_j) - \Phi(t_k - \mu_j)] / P_{jk}\}^2 \quad (67)$$

However

$$E[y_{jr}^2 | \mu_j, t_{k-1} < y < t_k] = \mu_j^2 + 2\mu_j[\phi(t_{k-1}-\mu_j) - \phi(t_k-\mu_j)]/P_{jk} \\ + \int_{t_{k-1}-\mu_j}^{t_k-\mu_j} \varepsilon^2 f(\varepsilon) d\varepsilon / P_{jk} \quad (68)$$

Using (68) in (67)

$$\text{Var}(y_{jr} | \mu_j, t_{k-1} < y < t_k) = \frac{1}{P_{jk}} \int_{t_{k-1}-\mu_j}^{t_k-\mu_j} \varepsilon^2 f(\varepsilon) d\varepsilon - \left[ \frac{\phi(t_{k-1}-\mu_j) - \phi(t_k-\mu_j)}{P_{jk}} \right]^2 \quad (69)$$

Completing the integration, (69) becomes

$$\text{Var}(y_{jr} | \mu_j, t_{k-1} < y < t_k) = \\ 1 + \frac{(t_{k-1}-\mu_j) \phi(t_{k-1}-\mu_j) - (t_k-\mu_j) \phi(t_k-\mu_j)}{P_{jk}} - \left[ \frac{\phi(t_{k-1}-\mu_j) - \phi(t_k-\mu_j)}{P_{jk}} \right]^2 \quad (70)$$

The meaning of (66) and (70) in relation to the method discussed in this paper becomes clear when a binary response is considered. Using the notation of the previous section

$$P_{j1} = \int_{-\infty}^{\gamma_j} \phi(y) dy ; P_{j2} = 1 - P_{j1}$$

where  $\phi(y)$  is a standard normal density. Then, using (66)

$$E(y|y > \gamma_j) = \frac{\phi(\gamma_j)}{P_{j2}} = i_{j2} \quad (71a)$$

$$E(y|y < \gamma_j) = -\frac{\phi(\gamma_j)}{P_{j1}} = i_{j1} \quad (71b)$$

Similarly, using (70)

$$\text{Var}(y|y > \gamma_j) = 1 - i_{j2}(i_{j2} - \gamma_j) \quad (72a)$$

$$\text{Var}(y|y < \gamma_j) = 1 - i_{j1}(i_{j1} - \gamma_j) \quad (72b)$$

In the binary case,  $v_j$  of (27) can be written as

$$v_j = n_{j1} i_{j1} + n_{j2} i_{j2} \quad (73)$$

so  $v_j$  is the sum of "normal scores" for the  $j^{\text{th}}$  row of the contingency table. Likewise,  $w_{jj}$  of (49) can be written as

$$w_{jj} = n_{j\cdot} \left[ \frac{\phi^2(\gamma_j)}{P_{j1}} + \frac{\phi^2(\gamma_j)}{P_{j2}} \right] \\ = n_{j\cdot} [P_{j1} i_{j1}^2 + P_{j2} i_{j2}^2] \quad (74)$$



which is the sum of variances of "normal scores" for the individuals in the  $j^{\text{th}}$  row. This is so because the mean "normal score" is  $P_{j1} i_{j1} + P_{j2} i_{j2} = 0$ .

Hence, equations (61) and (62) can be viewed as an algorithm to estimate, iteratively, the parameters of the model

$$\mathbf{K}\boldsymbol{\lambda} + \mathbf{H}\mathbf{h} + \mathbf{W}^{-1}\mathbf{v} \quad (75)$$

in which the  $\mathbf{v}$ 's are "normalized" residuals,

$$\text{Var} [\mathbf{K}\boldsymbol{\lambda} + \mathbf{H}\mathbf{h} + \mathbf{W}^{-1}\mathbf{v} | \boldsymbol{\lambda}, \mathbf{h}] = \mathbf{W}^{-1} \text{Var}(\mathbf{v}) \mathbf{W}^{-1} = \mathbf{W}^{-1} \quad (76)$$

and

$$\text{Var} [\mathbf{K}\boldsymbol{\lambda} + \mathbf{H}\mathbf{h} + \mathbf{W}^{-1}\mathbf{v} | \boldsymbol{\lambda}] = \mathbf{H}\mathbf{G}\mathbf{H}' + \mathbf{W}^{-1} \quad (77)$$

### VIII. Numerical example

An hypothetical data set used by SHAEFFER and WILTON (1976) in a discussion of sire evaluation for calving ease was employed to illustrate the procedures described in the present paper. The data consisted of calving ease scores from 28 male and female calves born in 2 herd-years from heifers and cows mated to 4 sires. Calving ease was scored as a response in one of 3 ordered categories (1 : normal birth ; 2 : slight difficulty, and 3 : extreme difficulty). The data, arranged into a  $20 \times 3$  contingency table are presented in Table 2. It should be emphasized that the records could have been arranged as a  $28 \times 3$  contingency table without changing the final results. As indicated in Table 3, about 68 % of the records were classified as easy calvings, 18 % as slightly difficult and about 14 % as decidedly difficult calvings. The four sires differed considerably in the distribution of calving ease scores. For example, while sires 1 and 4 had about the same proportion of easy calvings, they were markedly dissimilar with respect to the distribution of records in categories 2 and 3. However, and aside from sample size, the distribution of records by herd-years, cow age and sex were also different for these two sires illustrating the difficulties and pitfalls involved in ranking sires on the basis of raw frequencies.

*Model.* The model for the parameter  $\eta_j$  in (7), with  $j=1, \dots, 20$  indicating the  $j^{\text{th}}$  row of Table 3 was

$$\eta_j = H_k + A_\ell + S_m + u_n \quad (78)$$

where  $H_k$  is the effect of the  $k^{\text{th}}$  herd-year ( $k=1,2$ ),  $A_\ell$  is the effect of the  $\ell^{\text{th}}$  age of dam ( $\ell=1,2$  for heifers and cows, respectively),  $S_m$  is the effect of the  $m^{\text{th}}$  sex ( $m=1,2$  for males and females, respectively) and  $u_n$  is the effect of the  $n^{\text{th}}$  sire of calf ( $n=1, \dots, 4$ ). Then take

$$\mathbf{v}' = [H_1 + A_1 + S_1, H_2 - H_1, A_2 - A_1, S_2 - S_1] \quad (79)$$

from which  $\mathbf{Q}$  of order  $20 \times 4$  is easily obtained. It follows that

$$\mathbf{v} = H_1 + A_1 + S_1$$

and

$$\boldsymbol{\beta}' = [\beta_1, \beta_2, \beta_3] = [H_2 - H_1, A_2 - A_1, S_2 - S_1]$$

from which  $\mathbf{X}$  of order  $20 \times 3$  is easily deduced. The "working" variables to be estimated are then  $t_1 = \delta_1 - \mathbf{v}$ ,  $t_2 = \delta_2 - \mathbf{v}$ ,  $\boldsymbol{\beta}$ , and  $\mathbf{u}' = [u_1, u_2, u_3, u_4]$ .

TABLE 2

*Distribution of calving ease scores by herd-year, age of dam, sex of calf and sire of calf subclasses.*

*Répartition des données de facilité de vêlage selon le troupeau  $\times$  année, le sexe et le père du veau.*

Herd-year	Age of dam	Sex of calf	Sire of calf	Category of response <sup>a</sup>		
				1	2	3
1	2	M	1	1	0	0
1	2	F	1	1	0	0
1	3	M	1	1	0	0
1	2	F	2	0	1	0
1	3	M	2	1	0	1
1	3	F	2	3	0	0
1	2	M	3	1	1	0
1	3	F	3	0	1	0
1	3	M	3	1	0	0
2	2	F	1	2	0	0
2	2	M	1	1	0	0
2	3	M	1	0	0	1
2	2	F	2	1	0	1
2	3	M	2	1	0	0
2	2	F	3	0	1	0
2	3	M	3	0	0	1
2	2	M	4	0	1	0
2	2	F	4	1	0	0
2	3	F	4	2	0	0
2	3	M	4	2	0	0

a) 1 : normal birth; 2 : slight difficulty; 3 : extreme difficulty.

*Prior information.* Prior information about  $t_1$ ,  $t_2$  and  $\beta$  was assumed to be vague. The prior distribution for sire effects was taken as

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{I}/19)$$

The scalar 1/19 corresponds to an heritability of .20 in the underlying scale and to a residual variance equal to 1, i.e.,  $\sigma_e^2/\sigma_s^2 = (4-h^2)/h^2 = 19$ .

*Iteration.* Equations (52) were used for iteration. Different sets of starting values were employed and two such sets were :

$$\text{I) } t_k^{[0]} = \Phi^{-1} \left( \sum_{i=1}^k n_i / N \right), \quad k=1, 2,$$

where  $n_i$  is the total number of responses in category  $i$  and  $N$  is the total number of observations ;  $\hat{\beta}^{[0]} = \mathbf{0}$ , and  $\hat{\mathbf{u}}^{[0]} = \mathbf{0}$  ;

II) Same as above but with  $\mathbf{u}^{[0]}$  being a  $4 \times 1$  vector of random numbers generated as  $N(0, 1/19)$ .

TABLE 3

*Marginal relative frequencies of calving ease scores by level of the factors considered.*

*Fréquence marginale des facilités de vêlage par niveau de facteur.*

		Category of response <sup>a</sup>		
		1	2	3
Herd-year . . . . .	1	.6923	.2308	.0769
	2	.6667	.1333	.2000
Cow age . . . . .	2	.6154	.3077	.0769
	3	.7333	.0667	.2000
Sex of calf . . . . .	M	.6429	.1429	.2142
	F	.7143	.2143	.0714
Sire . . . . .	1	.8571	.0000	.1429
	2	.6667	.1100	.2222
	3	.3333	.5000	.1667
	4	.8334	.1666	.0000
Total . . . . .		.6786	.1786	.1428

a) 1 : normal birth; 2 : slight difficulty; 3 : extreme difficulty.

Iteration stopped when the average sum of squares of corrections was less than  $10^{-12}$ , i.e., when  $\Delta'/\Delta < 10^{-12}$ . In each of these two instances, convergence to the same solution occurred at the seventh round of iteration. The results of iteration beginning with Set I of starting values are presented in Table 4. From a practical point of view, iteration could have stopped at the fourth round if not earlier. Different sets of starting values for the thresholds were also examined. For example, 8 iterations were required for the sets  $\{t_1^{[0]} = \Phi^{-1}(0.85), t_2^{[0]} = \Phi^{-1}(0.95)\}$  and  $\{t_1^{[0]} = \Phi^{-1}(0.20), t_2^{[0]} = \Phi^{-1}(0.90)\}$ . It appears that the algorithm approaches the maximum rapidly from almost everywhere.

**Results.** Estimates of the parameters in the underlying scale and the square root of their posterior dispersion were :

$$\begin{aligned}
 t_1 &= .3753 \pm .5580 & \hat{S}_2 - \hat{S}_1 &= -.3907 \pm .4966 \\
 t_2 &= 1.0113 \pm .5789 & \hat{u}_1 &= -.0815 \pm .2139 \\
 t_2 - t_1 &= .6360 \pm .2565 & \hat{u}_2 &= .0655 \pm .2133 \\
 \hat{H}_2 - \hat{H}_1 &= .2975 \pm .4949 & \hat{u}_3 &= .1228 \pm .2147 \\
 \hat{A}_2 - \hat{A}_1 &= -.1269 \pm .4987 & \hat{u}_4 &= -.1068 \pm .2174
 \end{aligned}$$

TABLE 4

*Solutions by round of iteration with zero starting values for  $\hat{\beta}$  and  $\hat{\mu}$ .*

*Solutions aux différentes itérations en partant de valeurs initiales de  $\hat{\beta}$  et  $\hat{\mu}$  nulles.*

Parameter <sup>b)</sup>	Iteration <sup>a)</sup>				
	1	2	3	4	7
$t_1$	373.23	376.14	375.23	375.30	375.29
$t_2$	977.17	1010.32	1011.11	1011.35	1011.35
$\beta_1$	300.62	297.19	297.42	297.53	297.52
$\beta_2$	-124.65	-125.82	-126.91	-126.86	-126.87
$\beta_3$	-376.84	-389.32	-390.42	-390.65	-390.66
$\mu_1$	-80.02	-80.55	-81.50	-81.53	-81.54
$\mu_2$	63.25	64.69	65.43	65.50	65.50
$\mu_3$	119.93	121.31	122.75	122.79	122.80
$\mu_4$	-103.17	-105.45	-106.68	-106.76	-106.76

a) Convergences reached at round 7 with  $\Delta'\Delta/9 = 8 \times 10^{-15}$ .

b) Solutions given should be multiplied by  $10^{-3}$ .

In particular, it should be noted that the contrasts  $A_2-A_1$  (cows-heifers) and  $S_2-S_1$  (females-males) were estimated as negative, i.e., the probability of difficult calving for male calves and heifer calvings would be higher than for female calves and cows, respectively. This is in agreement with what one would expect from previous knowledge on the subject.

In animal breeding practice, the interest centers on estimation of response probabilities associated with specific linear combinations of parameters. For example, one may wish to calculate the probability of the event that a male calf out of the mating of the  $i^{\text{th}}$  bull to a heifer in herd 1, will experience a difficult birth (category 3). This is calculated as

$$1 - \Phi[\delta_2 - (\hat{H}_1 + \hat{A}_1 + \hat{S}_1 + \hat{u}_i)] = 1 - \Phi(t_2 - \hat{u}_i)$$

When this was applied to the four sires, the probability distribution by category of response was :

	Probability in category		
	1	2	3
Sire 1	.676	.187	.137
Sire 2	.622	.206	.172
Sire 3	.600	.213	.187
Sire 4	.685	.183	.132

Under artificial insemination, it is of interest to evaluate sires under more general conditions. For example, one may consider to estimate, for each sire, the probability distribution for heifer calvings across herds and sexes. For the first category, the four elementary probabilities would be

$$\Phi[\delta_1 - (H_k + A_1 + S_m + u_i)] \quad k=1,2 ; m=1,2 ; i = 1, \dots, 4$$

Then we can weigh these probabilities as

$$\sum_{k=1}^2 \sum_{m=1}^2 a_{km} \Phi[\delta_1 - (H_k + A_1 + S_m + u_i)]$$

such that  $\sum_k \sum_m a_{km} = a_{11} + a_{12} + a_{21} + a_{22} = 1$ . In the example and taking  $a_{11} = a_{12} = a_{21} = a_{22} = 1/4$ , this yields for the three categories.

	<i>Probability in category</i>		
	<i>1</i>	<i>2</i>	<i>3</i>
Sire 1	.688	.178	.134
Sire 2	.635	.197	.168
Sire 3	.614	.204	.182
Sire 4	.696	.175	.129

These estimates of probability are very different to the ones obtained on the basis of raw frequencies (Table 3) as they take into account the distribution of records across herd-years and sex of calf. Note that estimates are pulled towards the overall relative frequencies.

## IX. Conclusions

This paper describes a general non-linear procedure to analyze ordered categorical variates in the context of the data sets usually encountered in animal breeding practice. The model assumes an underlying continuous variable which is described as a linear combination of variables sampled from conceptual distributions. In contrast to other methods suggested for the analysis of categorical data, the procedure takes into account the assumption that candidates for selection are sampled from a distribution with known first and second moments, *a priori*. Theoretical problems arising when linear models are applied to categorical data (GIANOLA, 1982) are eliminated as the procedure adjusts automatically for differences in incidence among subpopulations considered in the analysis. In addition, the method can be further generalized to take into account the effect of concomitant variables, e.g., birth weight, on response probabilities. This will be reported in a future communication.

The estimation equations derived in this paper from a Bayesian viewpoint are equivalent to the extension of "generalized" linear models for all or none variables suggested by THOMPSON (1979), and to the approach presented by HARVILLE and MEE (1982) in a "maximum probability" estimation setting. Further, the method can be regarded as an algorithm to calculate "normal scores" as these appear naturally in the estimation equations.

In view of the computational requirements of the method, it is pertinent to address the question of how much better this non-linear predictor would be than standard linear model techniques (BERGER and FREEMAN, 1978). This is currently being examined via Monte-Carlo methods. However, the work of PORTNOY (1982) suggests that when predictands are intrinsically non-linear, predictors based on linear functions of the data may be poor for ranking purposes. Computational requirements could be decreased by speeding

up convergence by relaxation techniques, or by using other algorithms such as iterative proportional fitting in the binary case. HARVILLE and MEE (1982), in order to simplify computing, have suggested to change the coefficient matrix only after several rounds of iteration as opposed to every iterate.

The implementation of the procedure rests on the assumption that the variance-covariance matrix of the variables to be predicted is known. This assumption may be justified for some quantitative traits in animal breeding (e.g., milk yield) but not so for most categorical traits of economic importance as the methodology that has been used is questionable. It is theoretically possible to render the method completely Bayesian by taking a prior distribution for a covariance matrix (DE GROOT, 1970). This is a potentially interesting area for further work.

Although we have emphasized the normal distribution as a possible model for the underlying scale, the principles outlined in this paper permit working with alternative functional forms. For example, the probability of difficult calving could be expressed as

$$k \exp \{a+bx\} ; 0 < x < x_0 \\ 1 \qquad \qquad \qquad x \geq x_0$$

where  $x$  is a liability variable,  $a$  and  $b$  are functions of experimental conditions (age of dam, say), and  $k$  is a constant.

Received November 2, 1982.

Accepted January 31, 1983.

### Acknowledgements

Daniel GIANOLA wishes to acknowledge I.N.R.A., France, for support during his stay at Jouy-en-Josas, and the Holstein Association, Brattleboro, Vermont, U.S.A., for supporting his work in categorical data.

### References

- ASHTON W.D., 1972. *The Logit Transformation*. Hafner Publishing Co., New York.
- BERGER P.J., FREEMAN A.E., 1978. Prediction of sire merit for calving difficulty. *J. Dairy Sci.*, **61**, 1146-1150.
- BOCK R.D., 1975. *Multivariate Statistical Methods in Behavioral Research*. Mc Graw-Hill Book Co., New York.
- BOX G.E.P., TIAO G.C., 1973. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading.
- BULMER M.G., 1980. *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.
- COX D.R., HINKLEY D.V., 1974. *Theoretical Statistics*. Chapman and Hall, London.
- CURNOW R.N., SMITH C., 1975. Multifactorial models for familial diseases in man. *J.R. Statist. Soc. (A)*, **138**, 131-169.
- DAHLQUIST G. and BJORCK A., 1974. *Numerical Methods*. Prentice Hall, Englewood Cliffs.
- DE GROOT M.H., 1982. Comments on the role of parameters in the predictive approach to statistics. *Biometrics (Suppl.)*, **38**, 86-91.
- DE GROOT M.H., 1970. *Optimal Statistical Decisions*. Mc Graw-Hill, New York.

- DEMPSTER E.R., LERNER I.M., 1950. Heritability of threshold characters. *Genetics*, **35**, 212-235.
- FALCONER D.S., 1965. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann. Hum. Genet.*, **29**, 51-76.
- FALCONER D.S., 1967. The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitus. *Ann. Hum. Genet.*, **31**, 1-20.
- FERGUSON T.G., 1967. *Mathematical Statistics : A Decision Theoretic Approach*. Academic Press, New York.
- FERNANDO R.L., BILLINGSLEY R.D., GIANOLA D., 1983. Effects of method of scaling on heritability estimates and sire evaluations for frame size at weaning in Angus cattle. *J. Anim. Sci.* (In Press).
- GIANOLA D., 1982. Theory and analysis of threshold characters. *J. Anim. Sci.*, **54**, 1079-1096.
- GIANOLA D., FOULLEY J.L., 1982. Non-linear prediction of latent genetic liability with binary expression : an empirical Bayes approach. *Proc. 2nd Wld. Cong. Genet. Appl. Liv. Prod.*, **7**, 293-303, Madrid, Spain.
- GIANOLA D., NORTON H.W., 1981. Scaling threshold characters. *Genetics*, **99**, 357-364.
- HARVEY W.R., 1982. Least-squares analysis of discrete data. *J. Anim. Sci.*, **54**, 1079-1096.
- HARVILLE D.A., MEE R.W., 1982. *A mixed model procedure for analyzing ordered categorical data*. Mimeo, Dept. of Statistics, Iowa State University, Ames.
- HENDERSON C.R., 1973. Sire evaluation and genetic trends. In *Proc. of Animal Breeding and Genetics Symp.* in honor of Dr. L. L. Lush, Amer. Soc. of Anim. Sci. and Amer. Dairy Sci. Assoc., Champaign, Ill.
- HENDERSON C.R., 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423-447.
- HENDERSON C.R., 1977. Best linear unbiased prediction of breeding values not in the model for records. *J. Dairy Sci.*, **60**, 783-787.
- JOHNSON N.L., KOTZ S., 1970. *Distribution in Statistics*. Vol. 2, Continuous Univariate Distributions. Houghton Mifflin Co., Boston.
- KENDALL M.G., STUART A., 1961. *The Advanced Theory of Statistics*, Vol. 2, Hafner, New York.
- LEONARD T., 1972. Bayesian methods for binomial data. *Biometrika*, **59**, 581-589.
- LINDLEY D.V., 1965. *Introduction to Probability and Statistics from a Bayesian viewpoint*. Part 2 : Inference. Cambridge University Press, Cambridge.
- NELDER J.A., WEDDERBURN R.W.M., 1972. Generalized linear models. *J.R. Statist. Soc. A*, **135**, 370-384.
- PORTNOY S., 1982. Maximizing the probability of correctly ordering random variables using linear predictors. *J. Mult. Anal.*, **12**, 256-269.
- PRATT J.W., RAIFFA H., SCHLAIFER R., 1965. *Introduction to Statistical Decision Theory*. McGraw-Hill, New York.
- SCHAEFFER L.R., WILTON J.W., 1976. Methods of sire evaluation for calving ease. *J. Dairy Sci.*, **59**, 544-551.
- SEARLE S.R., 1971. *Linear Models*. Wiley, New York.
- SNELL E.J., 1964. A scaling procedure for ordered categorical data. *Biometrics*, **20**, 592-607.
- THOMPSON R., 1979. Sire evaluation. *Biometrics*, **35**, 339-353.
- TONG A.K.W., WILTON J.W., SCHAEFFER L.R., 1977. Application of a scoring procedure and transformations to dairy type classification and beef ease of calving categorical data. *Can. J. anim. Sci.*, **57**, 1-9.